EXHIBIT A

1: Kim SK, Wang JC. Related Articles, Links

Gene silencing via protein-mediated subcellular localization of DNA.
Proc Natl Acad Sci U S A. 1999 Jul 20;96(15):8557-61.
PMID: 10411914 [PubMed - indexed for MEDLINE]

2: Lum PL, Schildbach JF. Related Articles, Links

Specific DNA recognition by F Factor TraY involves beta-sheet residues.
J Biol Chem. 1999 Jul 9;274(28):19644-8.
PMID: 10391902 [PubMed - indexed for MEDLINE]

3: Radkov SA, Touitou R, Brehm A, Rowe M, West M, Kouzarides T, Allday MJ. Related Articles, Links

Epstein-Barr virus nuclear antigen 3C interacts with histone deacetylase to repress transcription.
J Virol. 1999 Jul;73(7):5688-97.
PMID: 10364319 [PubMed - indexed for MEDLINE]

4: Abu-Elneel K, Kapeller I, Shlomai J. Related Articles, Links

Universal minicircle sequence-binding protein, a sequence-specific DNA-binding protein that recognizes the two replication origins of the kinetoplast DNA minicircle.
J Biol Chem. 1999 May 7;274(19):13419-26.
PMID: 10224106 [PubMed - indexed for MEDLINE]

5: Ehrlich KC, Montalbano BG, Cary JW. Related Articles, Links

Binding of the C6-zinc cluster protein, AFLR, to the promoters of aflatoxin pathway biosynthesis genes in Aspergillus parasiticus.
Gene. 1999 Apr 16;230(2):249-57.
PMID: 10216264 [PubMed - indexed for MEDLINE]

6: Bartsch O, Horstmann S, Toprak K, Klempnauer KH, Ferrari S. Related Articles, Links

Identification of cyclin A/Cdk2 phosphorylation sites in B-Myb.
Eur J Biochem. 1999 Mar;260(2):384-91.
PMID: 10095772 [PubMed - indexed for MEDLINE]

7: Von Ohlen T, Hooper JE. Related Articles, Links

The ciD mutation encodes a chimeric protein whose activity is regulated by Wingless signaling.
Dev Biol. 1999 Apr 1;208(1):147-56.

PMID: 10075848 [PubMed - indexed for MEDLINE]

8: Tolnay M, Vereshchagina LA, Tsokos GC. Related Articles, Links

Heterogeneous nuclear ribonucleoprotein D0B is a sequence-specific DNA-binding protein.

Biochem J. 1999 Mar 1;338 (Pt 2):417-25.

PMID: 10024518 [PubMed - indexed for MEDLINE]

9: Walton M, Saura J, Young D, MacGibbon G, Hansen W, Lawlor P, Sirimanne E, Gluckman P, Dragunow M. Related Articles, Links

CCAAT-enhancer binding protein alpha is expressed in activated microglial cells after brain injury.

Brain Res Mol Brain Res. 1998 Oct 30;61(1-2):11-22.

PMID: 9795105 [PubMed - indexed for MEDLINE]

10: Banecki B, Kaguni JM, Marszalek J. Related Articles, Links

Role of adenine nucleotides, molecular chaperones and chaperonins in stabilization of DnaA initiator protein of Escherichia coli.

Biochim Biophys Acta. 1998 Oct 23;1442(1):39-48.

PMID: 9767098 [PubMed - indexed for MEDLINE]

11: Sharpe PL, Craig NL. Related Articles, Links

Host proteins can stimulate Tn7 transposition: a novel role for the ribosomal protein L29 and the acyl carrier protein.

EMBO J. 1998 Oct 1;17(19):5822-31.

PMID: 9755182 [PubMed - indexed for MEDLINE]

12: Simonsson S, Samuelsson T, Elias P. Related Articles, Links

The herpes simplex virus type 1 origin binding protein. Specific recognition of phosphates and methyl groups defines the interacting surface for a monomeric DNA binding domain in the major groove of DNA.

J Biol Chem. 1998 Sep 18;273(38):24633-9.

PMID: 9733759 [PubMed - indexed for MEDLINE]

13: Kortschak RD, Reimann H, Zimmer M, Eyre HJ, Saint R, Jenne DE. Related Articles, Links

The human dead ringer/bright homolog, DRIL1: cDNA cloning, gene structure, and mapping to D19S886, a marker on 19p13.3 that is strictly linked to the Peutz-Jeghers syndrome.

Genomics. 1998 Jul 15;51(2):288-92.

PMID: 9722953 [PubMed - indexed for MEDLINE]

14: Raina R, Schlappi M, Karunanandaa B, Elhofy A, Fedoroff N. Related Articles, Links

Concerted formation of macromolecular Suppressor-mutator transposition complexes.

Proc Natl Acad Sci U S A. 1998 Jul 21;95(15):8526-31.

PMID: 9671711 [PubMed - indexed for MEDLINE]

15: Brown JL, Mucci D, Whiteley M, Dirksen ML, Kassis JA. Related Articles, Links

The Drosophila Polycomb group gene pleiohomeotic encodes a DNA binding protein with homology to the transcription factor YY1.

Mol Cell. 1998 Jun;1(7):1057-64.

PMID: 9651589 [PubMed - indexed for MEDLINE]

16: Ariumi Y, Shimotohno K, Noda M, Hatanaka M. Related Articles, Links
Characterization of the internal promoter of human T-cell leukemia virus type I.
FEBS Lett. 1998 Feb 13;423(1):25-30.
PMID: 9506835 [PubMed - indexed for MEDLINE]

17: Boehmer PE. Related Articles, Links
The herpes simplex virus type-1 single-strand DNA-binding protein, ICP8, increases the processivity of the UL9 protein DNA helicase.
J Biol Chem. 1998 Jan 30;273(5):2676-83.
PMID: 9446572 [PubMed - indexed for MEDLINE]

18: Sutton MD, Kaguni JM. Related Articles, Links
The Escherichia coli dnaA gene: four functional domains.
J Mol Biol. 1997 Dec 12;274(4):546-61.
PMID: 9417934 [PubMed - indexed for MEDLINE]

19: Kitada T, Seki S, Nakatani K, Kawada N, Kuroki T, Monna T. Related Articles, Links
Hepatic expression of c-Myb in chronic human liver disease.
Hepatology. 1997 Dec;26(6):1506-12.
PMID: 9397991 [PubMed - indexed for MEDLINE]

20: Gariglio M, Ying GG, Hertel L, Gaboli M, Clerc RG, Landolfo S. Related Articles, Links
The high-mobility group protein T160 binds to both linear and cruciform DNA and mediates DNA bending as determined by ring closure.
Exp Cell Res. 1997 Nov 1;236(2):472-81.
PMID: 9367632 [PubMed - indexed for MEDLINE]

Display Summary Show: 20 Sort Send to Text

Items 1-20 of 289 Page 1 of 15 Next

Write to the Help Desk
NCBI | NLM | NIH
Department of Health & Human Services
Freedom of Information Act | Disclaimer

EXHIBIT B

Limits: **Publication Date to 1999/08/26**

1: Gaszner M, Vazquez J, Schedl P. Related Articles, Links

The Zw5 protein, a component of the scs chromatin domain boundary, is able to block enhancer-promoter interaction.
Genes Dev. 1999 Aug 15;13(16):2098-107.
PMID: 10465787 [PubMed - indexed for MEDLINE]

2: Dhavan GM, Lapham J, Yang S, Crothers DM. Related Articles, Links

Decreased imino proton exchange and base-pair opening in the IHF-DNA complex measured by NMR.
J Mol Biol. 1999 May 14;288(4):659-71.
PMID: 10329171 [PubMed - indexed for MEDLINE]

3: Baillie RA, Sha X, Thuillier P, Clarke SD. Related Articles, Links

A novel 3T3-L1 preadipocyte variant that expresses PPARgamma2 and RXRalpha but does not undergo differentiation.
J Lipid Res. 1998 Oct;39(10):2048-53.
PMID: 9788251 [PubMed - indexed for MEDLINE]

4: Simonsson S, Samuelsson T, Elias P. Related Articles, Links

The herpes simplex virus type 1 origin binding protein. Specific recognition of phosphates and methyl groups defines the interacting surface for a monomeric DNA binding domain in the major groove of DNA.
J Biol Chem. 1998 Sep 18;273(38):24633-9.
PMID: 9733759 [PubMed - indexed for MEDLINE]

5: Abidi FE, Roh H, Keath EJ. Related Articles, Links

Identification and characterization of a phase-specific, nuclear DNA binding protein from the dimorphic pathogenic fungus Histoplasma capsulatum.
Infect Immun. 1998 Aug;66(8):3867-73.
PMID: 9673274 [PubMed - indexed for MEDLINE]

6: Cox GS, Gutkin DW, Haas MJ, Cosgrove DE. Related Articles, Links

Isolation of an Alu repetitive DNA binding protein and effect of CpG methylation on binding to its recognition sequence.
Biochim Biophys Acta. 1998 Mar 4;1396(1):67-87.
PMID: 9524225 [PubMed - indexed for MEDLINE]

7: Sun W, Hattman S, Kool E. Related Articles, Links

Interaction of the bacteriophage Mu transcriptional activator protein, C, with

its target site in the mom promoter.
J Mol Biol. 1997 Nov 7;273(4):765-74.
PMID: 9367769 [PubMed - indexed for MEDLINE]

8: Nambiar A, Swamynathan SK, Kandala JC, Guntaka RV. Related Articles, Links
Characterization of the DNA-binding domain of the avian Y-box protein, chkYB-2, and mutational analysis of its single-strand binding motif in the Rous sarcoma virus enhancer.
J Virol. 1998 Feb;72(2):900-9.
PMID: 9444981 [PubMed - indexed for MEDLINE]

9: Coupe SA, Deikman J. Related Articles, Links
Characterization of a DNA-binding protein that interacts with 5' flanking regions of two fruit-ripening genes.
Plant J. 1997 Jun;11(6):1207-18.
PMID: 9225464 [PubMed - indexed for MEDLINE]

10: Keren-Tal I, Dantes A, Plehn-Dujowich D, Amsterdam A. Related Articles, Links
Association of Ad4BP/SF-1 transcription factor with steroidogenic activity in oncogene-transformed granulosa cells.
Mol Cell Endocrinol. 1997 Mar 14;127(1):49-57.
PMID: 9099900 [PubMed - indexed for MEDLINE]

11: Sawada Y, Noda M. Related Articles, Links
An adipogenic basic helix-loop-helix-leucine zipper type transcription factor (ADD1) mRNA is expressed and regulated by retinoic acid in osteoblastic cells.
Mol Endocrinol. 1996 Oct;10(10):1238-48.
PMID: 9121491 [PubMed - indexed for MEDLINE]

12: Martin SF, Spaller MR, Hergenrother PJ. Related Articles, Links
Expression and site-directed mutagenesis of the phosphatidylcholine-preferring phospholipase C of Bacillus cereus: probing the role of the active site Glu146.
Biochemistry. 1996 Oct 1;35(39):12970-7.
PMID: 8841144 [PubMed - indexed for MEDLINE]

13: Blake M, Niklinski J, Zajac-Kaye M. Related Articles, Links
Interactions of the transcription factors MIBP1 and RFX1 with the EP element of the hepatitis B virus enhancer.
J Virol. 1996 Sep;70(9):6060-6.
PMID: 8709229 [PubMed - indexed for MEDLINE]

14: Thomas M, Massimi P, Banks L. Related Articles, Links
HPV-18 E6 inhibits p53 DNA binding activity regardless of the oligomeric state of p53 or the exact p53 recognition sequence.
Oncogene. 1996 Aug 1;13(3):471-80.
PMID: 8760288 [PubMed - indexed for MEDLINE]

15: Liu PC, Phillips MA, Matsumura F. Related Articles, Links
Alteration by 2,3,7,8-Tetrachlorodibenzo-p-dioxin of CCAAT/enhancer binding protein correlates with suppression of adipocyte differentiation in

3T3-L1 cells.
Mol Pharmacol. 1996 Jun;49(6):989-97.
PMID: 8649359 [PubMed - indexed for MEDLINE]

16: Chang L, Thompson MA. Related Articles, Links
Activity of the distal positive element of the peripherin gene is dependent on proteins binding to an Ets-like recognition site and a novel inverted repeat site.
J Biol Chem. 1996 Mar 15;271(11):6467-75.
PMID: 8626448 [PubMed - indexed for MEDLINE]

17: Chen Y, Gill GN. Related Articles, Links
A heterodimeric nuclear protein complex binds two palindromic sequences in the proximal enhancer of the human erbB-2 gene.
J Biol Chem. 1996 Mar 1;271(9):5183-8.
PMID: 8617800 [PubMed - indexed for MEDLINE]

18: Walker GT, Linn CP, Nadeau JG. Related Articles, Links
DNA detection by strand displacement amplification and fluorescence polarization with signal enhancement using a DNA binding protein.
Nucleic Acids Res. 1996 Jan 15;24(2):348-53.
PMID: 8628661 [PubMed - indexed for MEDLINE]

19: Lian JB, Stein GS, Stein JL, Van Wijnen A, McCabe L, Banerjee C, Hoffmann H. Related Articles, Links
The osteocalcin gene promoter provides a molecular blueprint for regulatory mechanisms controlling bone tissue formation: role of transcription factors involved in development.
Connect Tissue Res. 1996;35(1-4):15-21.
PMID: 9084639 [PubMed - indexed for MEDLINE]

20: Samadani U, Qian X, Costa RH. Related Articles, Links
Identification of a transthyretin enhancer site that selectively binds the hepatocyte nuclear factor-3 beta isoform.
Gene Expr. 1996;6(1):23-33.
PMID: 8931989 [PubMed - indexed for MEDLINE]

Display Summary Show: 20 Sort Send to Text

Items 1-20 of 54 Page 1 of 3 Next

Write to the Help Desk
NCBI | NLM | NIH
Department of Health & Human Services
Freedom of Information Act | Disclaimer

i686-pc-linux-gnu Jan 7 2003 16:40:32

EXHIBIT C

Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 12357–12361, December 1994
Biophysics

DNA recognition code of transcription factors in the helix–turn–helix, probe helix, hormone receptor, and zinc finger families

(DNA–protein interaction/homeodomain/leucine zipper/transcription factor GATA)

MASASHI SUZUKI* AND NAOTO YAGI†

*Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, United Kingdom; and †Tohoku University, School of Medicine, Seiryo-machi, Sendai, 980-77, Japan

Communicated by Tadamitsu Kishimoto, August 8, 1994

ABSTRACT **We have previously reported that in four transcription factor families the DNA-recognition rules can be described as (*i*) chemical rules, which list possible pairings between the 20 amino acid residues and the four DNA bases, and (*ii*) stereochemical rules, which describe the base and amino acid positions in contact. We have incorporated these rules into a computer program and examined the nature of the rules. Here we conclude that the DNA recognition rules are simple, logical, and consistent. The rules are specific enough to predict DNA-binding characteristics from a protein sequence.**

A large number of transcription factors, which play dominant roles in transcription regulation by binding to different DNA sequences, have been identified. Since the three-dimensional structure of a protein is uniquely fixed by its amino acid sequence, basic rules are expected, which would predict the DNA-binding specificity from a transcription factor sequence. But, since the initial expectation of such rules (the recognition code) (1), many structural biologists have expressed skepticism about their existence (for example, see ref. 2).

The crystal structures of a number of transcription factor–DNA complexes have been determined (3–27); also a considerable amount of biochemical, genetic, and statistical information about the binding specificity of transcription factors is available (28–34). By using these data, we have devised a method of analyzing the patterns of contacts between DNA bases and amino acid residues (35–40) and have described the DNA-recognition rules of four transcription factor families: the probe helix (PH), which includes homeo and zipper proteins (35, 36); the helix–turn–helix (HTH) (M.S. and M. Gerstein, unpublished results); the zinc finger (ZnF) (37, 38); and the C4 Zn-binding proteins (C4), which include hormone receptors and GATA proteins (38–40). These rules concern contacts from amino acid side chains in a recognition helix to DNA bases in the major groove.

The aim of this paper is to establish a framework of DNA-recognition rules common to the four families and to examine whether, from the nature of the rules, they constitute a recognition code.

Framework of the DNA Recognition Rules

The DNA-recognition rules are of two types, chemical and stereochemical. The chemical rules list possible pairing partners of amino acid side chains and DNA bases through hydrogen bonding or hydrophobic interaction (Fig. 1*a*; ref. 36). The sizes of residues are also important; from a fixed position on an interaction surface, a longer side chain can reach a more distant part of the DNA. The residues are classified roughly into four groups—small, medium, large, and aromatic (Fig. 1*a*; ref. 36). These chemical rules are *general* for any binding motif.

The inclination of the recognition helix in the major groove of DNA is fixed by the structural elements specific to a DNA-binding motif. For instance, a recognition helix of PH has conserved Arg/Lys positions, which bind to DNA phosphates and thereby fix the binding geometry (35, 36). As a consequence, each binding motif uses a set of particular amino acid positions for base recognition. These can be easily summarized into a chart with specifications of the sizes of residues used; each DNA-binding motif has its own *specific* stereochemical chart (Fig. 1 *b–e*). ZnF motifs can be subdivided into two groups (37), but here only the larger group is discussed (A fingers).

Binding Score

We have incorporated the rules into a computer program, which is written in the C programming language and implemented under the Unix operating system. Its core function is to score the match between the given DNA and protein sequences. This binding score is essentially the number of contacts predicted between the two sequences and reflects the binding energy.

To calculate the binding score, points for stereochemical (see the legend to Fig. 1 *b–e*) and chemical (Fig. 1*a*) merits are introduced. The binding score is calculated as the sum over all the contacts of (stereochemical merit point) × (chemical merit point) for each interaction. The chemical merit points given to different base–residue partners are not always the same (Fig. 1*a*). For instance, Arg and Lys could bind by a hydrogen bond to T, G, or A. But in fact they recognize the G base almost exclusively (36), because the G base in a G·C pair is electrically polar (negatively charged), while Arg and Lys have a positive charge. Therefore, binding of Arg or Lys to G should be given more points than to T or A. Similarly, not all the contacts in the stereochemical charts appear to be equally important (refs. 36 and 37; M.S. and M. Gerstein, unpublished results), and this is reflected in differences in the two grades of stereochemical merit points (see contacts marked with diamonds and those not in Fig. 1 *b–e*).

Often several different sets of contacts are possible for given protein and DNA sequences. In this case, the pairing with the highest score is chosen. However, it is stereochemically forbidden to make two contacts that cross over each other in the chart. For instance, in Fig. 1*c* aa 5 can contact C3, and aa 8 can contact C2, but not simultaneously. As an example, the binding score of CAP (Fig. 2*h*) is the sum of the products of the chemical and stereochemical points for

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: PH, probe helix; HTH, helix–turn–helix; ZnF, zinc finger; C4, C4 Zn-binding protein.

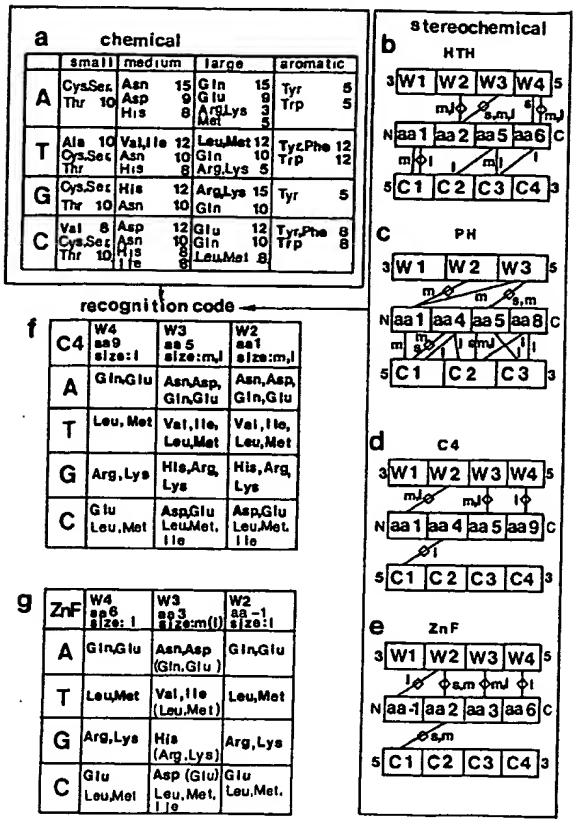

FIG. 1. Chemical (*a*) and stereochemical (*b–e*) rules that make the DNA-recognition code and code tables for C4 (*f*) and ZnF (*g*). (*a*) The chemical merit points are also shown. Residues in boldfaced letters are those important for specificity (specific residues). (*b–e*) Sketches of the DNA major groove with the bases, W1–W4 (top strand) and C1–C4 (bottom strand), to which a recognition helix (in the middle line) binds. The sizes of residues, small (s), medium (m), and large (l), used for the contacts are also shown. Aromatic residues may often be included with the large group. Ten stereochemical merit points are given to the contacts marked with diamonds and five to the other contacts. No stereochemical points are given otherwise. If a hydrophobic interaction takes place to a T base and if one of the two neighboring bases is another T, an additional 3 points is added to the chemical merit point, since this is likely to enhance the hydrophobic environment. The binding specificity of Asn (aa 1) of PH is affected by Asn (aa 2) through side chain–side chain interactions (36); if Asn occupies position 2, Asn (aa 1) interacts with Asn (aa 2) and binds to A (W2), but if not Asn (aa 2) bridges the C1 and W2 bases. For this reason, if position 2 is occupied by Asn, the chemical merit point of Asn (aa 1) to A (W2) is kept at 15; if not, it is decreased to 10 and the residue is allowed to bind to the C1 base at the same time. When a single residue binds to two bases simultaneously, the two contacts are handled independently. This is to simplify the computer program, although the two bases bridged in this way are limited and can be handled as a set (36). The code tables (*f* and *g*) are made by choosing the columns from *a* according to the residue sizes specified in *d* and *e*. The interaction of hydrophobic residues to the C base is weaker and therefore is shown by plain instead of boldfaced letters. Position 3 in ZnF can be occupied by a medium or large residue, but a medium residue is preferable (37); the large residues are shown in the parentheses.

the Arg·G, Arg·G, and Glu·C contacts, respectively—(10 × 15) + (5 × 15) + (10 × 12) = 345.

Consistency and Specificity of the Rules

The DNA recognition rules were originally deduced from 25 crystal structures (3–27) and many other transcription factors whose binding specificity has been characterized by genetic or biochemical experiments (see the references cited in refs. 35–40).

Contacts were predicted by the program for 73 recognition helices: those of 10 PH proteins, 20 HTH proteins, 38 ZnF proteins (specific or very specific A fingers listed in ref. 37), and 5 C4 proteins (selected examples are shown in Fig. 2).

In most examples, the predicted contacts are essentially the same as those observed or predicted in earlier work. Thus the rules can consistently explain the amino acid–base contacts. However, this does not necessarily suggest that the rules can explain how factors discriminate between the target and other DNA sequences; if many other DNA sequences were recognized by a factor in similar ways, the factor could not choose the correct site. We now examine this aspect (specificity) of the rules in two ways.

We first compare the binding score given to the real binding site with those for sites consisting of all other possible base combinations (Fig. 3). HTH, C4, and ZnF recognize four base pairs, which have 256 possible combinations. PH recognizes three base pairs, and the number of combinations is 64. In our calculation, the real binding sequence is usually found among a small number of DNA sequences that score the highest (Fig. 3); the rules are sufficiently specific to exclude the rest of the DNA sequences, which score less. To evaluate the specificity of the rules, we introduce the specificity index, which is defined as $(100 - n - \frac{m}{2})\%$, where n is the percentage of the DNA sequences that score higher than the real binding sequence and m is that of the DNA sequences that score the same as the real binding sequence. If a factor has two natural binding sequences—sequence i, which scores higher than sequence j—n is defined as the percentage that scores higher than i, and m is defined as the percentage that scores between i and j. The average indices calculated for the factors are 93% (PH) (96% if Max is excluded, which is further discussed in M.S. and M. Gerstein, unpublished results), 99% (C4), 96% (ZnF), and 92% (HTH).

As a second test we now examine the DNA sequence of a region regulated by a transcription factor *in vivo*. When the binding score is calculated for every four base pairs along the DNA, shifting one base pair at a time, the highest score is given for the experimentally identified binding site (Fig. 4). Since DNA has two strands, the score must be calculated along each of the two strands.

The above two tests have shown that the rules are highly specific. In the crystal structures, some additional contacts are seen from outside a recognition helix, but the binding specificity of a recognition helix seems to be essentially sufficient to specify uniquely the DNA-binding sites.

Spacing Type

An α-helix can bind to no more than five base pairs because of the curvature of the DNA major groove; it can access only one side of the DNA (44). To recognize more than five base pairs, two or more helices are used in combination, essentially by either relating the two with a twofold symmetry axis or repeating them in tandem. The classic HTH proteins and zipper proteins of the PH family use "symmetrical" arrangements (denoted here as S), while ZnF proteins use a "tandem" arrangement (denoted here as T). C4 proteins use both types of arrangements (45).

Symmetrical arrangements can be characterized by whether the C terminus (denoted with the "+" sign) or the N terminus (denoted with the "−" sign) is closer to the dyad axis and the number of bases along the DNA between the two binding sites (for example, S +6 for the HTH protein CAP). By knowing the spacing type, the plot of the binding score can be improved. When the binding scores of the two DNA strands for CAP binding are shifted by six base pairs and added to each other, the

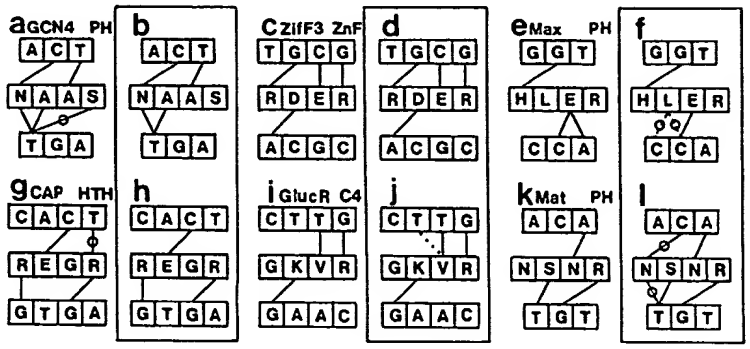

FIG. 2. Comparison between contacts observed in the crystal structures (*a*, *c*, *e*, *g*, *i*, and *k*) and computer-predicted contacts (*b*, *d*, *f*, *h*, *j*, and *l*). The figures are drawn in the same way as in Fig. 1. The dotted line (· · ·) in *j* shows an additional predicted hydrophobic interaction to the neighboring T base. A pair of two dashed lines (– – –) in *f* show two alternative contacts with the same score. The contacts that are predicted but not observed and those observed but not predicted are marked with circles. The side chain of Asn (aa 1) in Matα2 (*k*) is not well described in the original report of the crystal structure (4). The residue is predicted to contact the C (C1) and T (W2) bases (*l*). Leu (aa 4) of Max is predicted to make contacts with C (C1) or C (C2) (*f*). The figures of the original report (5) show that this leucine does seem close to C (C1), but the coordinates have not been published and the paper does not mention this contact.

new plot shows a clearer peak (Fig. 4*e*). Thus, a weaker binding specificity of a HTH recognition helix (see the previous section) is compensated by combining two such helices.

The spacing type of the majority of ZnF proteins is T −1 [i.e., two neighboring fingers share one base pair (−1) in a tandem (T) arrangement (37)]. A single finger appears to be incapable of discriminating between DNA sequences, but the combination of two or three fingers does seem to be sufficient (see figure 9 of ref. 37). This can explain why fingers are always found in a repeat.

The two experimentally identified ADR1 (ZnF)-binding sites in its regulatory DNA region are predicted successfully (Fig. 4*c*). The two sites are likely to be recognized by a symmetrical dimer of ADR1 molecules, each of which has two ZnF motifs in tandem (T −1), with the superspacing type of S +6 (Fig. 4*c*). Therefore, the communication between DNA and proteins can be described with increasing accuracy, from the chemical, the stereochemical, the spacing to the superspacing levels.

Prediction and Design

Our computer program successfully identifies the binding sites of transcription factors whose binding specificities have been characterized experimentally. Therefore, it may be natural to expect that it can (*i*) predict the yet unknown binding specificity of a protein sequence and (*ii*) design a factor that would recognize a particular DNA sequence.

In the ZnF and C4 families, a simple table relating DNA and protein sequences can be produced (Fig. 1 *f* and *g*; ref. 38). Three residues of C4—1, 5, and 9—bind to the three consecutive bases W2–W4, by a simple one residue–one base relationship, while ZnF positions −1, 3, and 6 bind to W2–W4. Therefore, by choosing specific partner residues in the correct columns from Fig. 1*a* according to the amino acid sizes shown in Fig. 1 *d* and *e*, recognition tables for the three positions from two types can be constructed (see ref. 38 for further discussion).

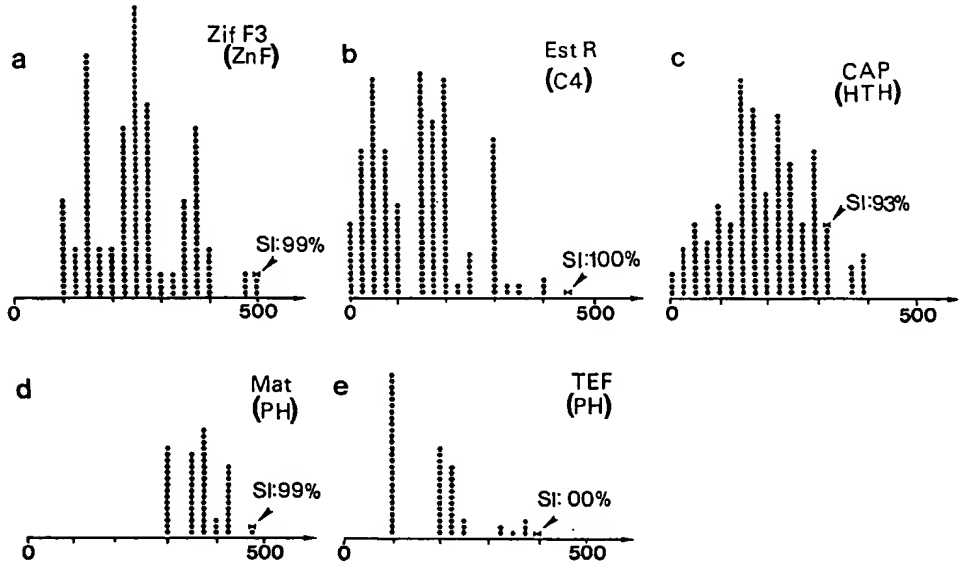

FIG. 3. Distribution of the binding scores for Zif268 finger 3 (ZnF) (*a*), estrogen receptor (C4) (*b*), CAP (HTH) (*c*), Matα2 (PH) (*d*), and TEF (PH) (*e*). The scores given to the real binding sites (marked with arrowheads) are compared with those given to the rest of all the possible combinations of DNA bases. The abscissas show the binding score, while the ordinates show the number of DNA sequences with that score. The specificity index (SI) is also shown. Note that TEF has Asn (aa 1) and Asn (aa 2) but Matα2 has Asn only at position 1 (see legend to Fig. 1).

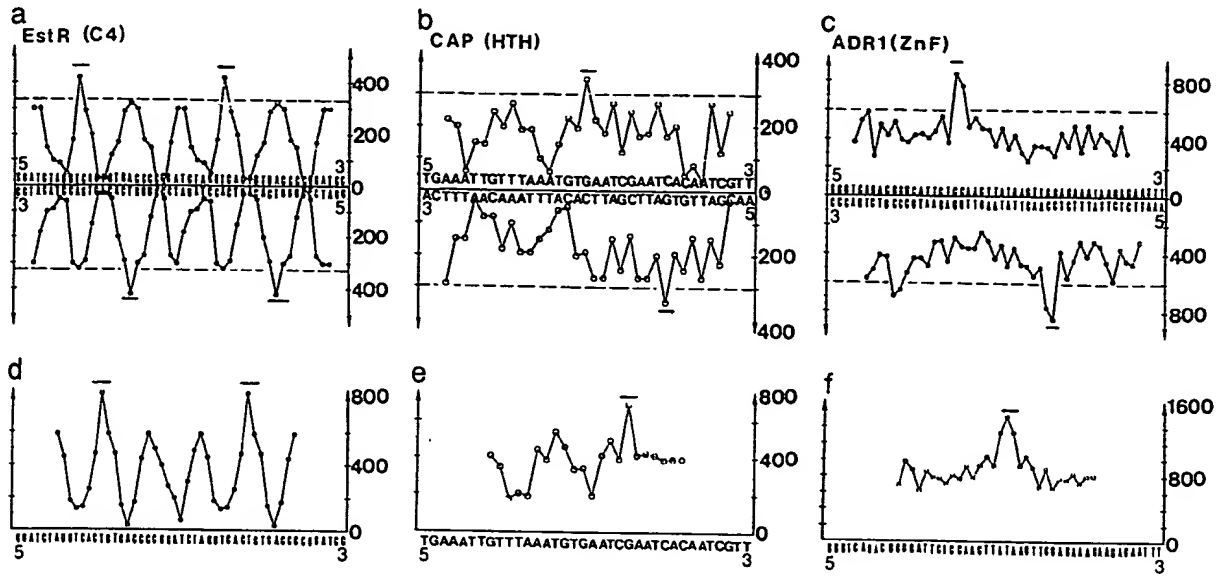

FIG. 4. Prediction of the binding sites for factors: estrogen receptor (C4) (*a* and *d*), CAP (HTH) (*b* and *e*), and ADR1 (ZnF) (*c* and *f*). (*a–c*) The binding score is calculated at every four base pairs by shifting one base pair along the DNA strand at a time. The DNA sequences were taken from refs. 41–43. The experimentally identified binding sites are marked with bars. The dotted lines show the cut-off levels, which separate real peaks from the background. (*d–f*) The binding scores to the two DNA strands are added to each other according to the spacing types. Note that a new peak for a dimer turns up in the center of two monomer binding sites on different DNA strands. The spacing types of symmetrical arrangements identified are S +1, PH (HDZip, bZip); S +2, PH (bZip, bHLH) and C4 (ThyR); S −2, HTH (RafR); S −3, HTH (EbgR, MalR) S −4, HTH (LacR, GalR); S +5, C4 (EstR, GlcR; although the three base pairs at the center of the binding sites are often described as the spacer, because these sequences vary, here the five base pairs that are not contacted by the recognition helices are defined as the spacer); S −6, HTH (DeoR) and C6 (PPR1); S +6, HTH (CAP, 434C, 434R, 16-3R); S +7, HTH (λR, λC); S +8, TrpR; S +8, HTH (CytR); S −10, HTH (P22C, P22R, LexA) and C6 (PUT3), S −11, C6 (Gal4). The spacing types of tandem arrangements identified are T −1, ZnF(A)–ZnF(A); T 0,ZnF(B)–ZnF(B); T +1, ZnF(B)–ZnF(A); T +3, C4(RXR)–C4(RAR), C4(RXR)–C4(COUP), C4(RXR)–C4(PPAR), C4(RXR)–C4(RXR); T +5, C4(RXR)–C4(VDR); T +6, C4(RXR)–C4(ThyR); T +7, C4(RXR)–C4(RAR).

The rules will be further improved as information becomes available. For example, in this study, changes in the DNA structure upon binding proteins and the sequence-dependent differences in the DNA structures are ignored. However, the framework and the major features of the rules are unlikely to change. We have shown that the DNA-recognition rules for well-characterized factors in the four families are simple, logical, consistent, and specific. We therefore believe that these rules constitute the DNA-recognition code.

We thank Drs. C. Chothia, J. Finch, and A. Klug and Mr. S. E. Brenner for their critical reading of the paper.

1. Pabo, C. O. & Sauer, R. T. (1984) *Annu. Rev. Biochem.* **53**, 293–321.
2. Matthews, B. W. (1988) *Nature (London)* **335**, 294–295.
3. Pabo, C. O., Aggarwal, A. K., Jordan, S. R., Beamer, L. J., Obeysekare, U. R. & Harrison, S. C. (1990) *Science* **247**, 1210–1213.
4. Wolberger, C., Vershon, A. K., Liu, B., Johnson, A. D. & Pabo, C. O. (1991) *Cell* **67**, 517–528.
5. Ferré-D'Amaré, A. R., Prendergast, G. C., Ziff, E. B. & Burley, S. K. (1993) *Nature (London)* **363**, 38–45.
6. Ellenberger, T. E., Brandl, C. S., Struhl, K. & Harrison, S. C. (1992) *Cell* **71**, 1223–1237.
7. König, P. & Richmond, T. (1993) *J. Mol. Biol.* **233**, 139–154.
8. Ferré-D'Amaré, A. R., Pognonec, P., Roeder, R. G. & Burley, S. K. (1994) *EMBO J.* **13**, 180–189.
9. Clarke, N. D., Beamer, L. J., Goldberg, H. R., Berkower, C. & Pabo, C. O. (1991) *Science* **254**, 267–270.
10. Schwabe, J. W., Chapman, L., Finch, J. T. & Rhodes, D. (1993) *Cell* **75**, 567–578.
11. Omichinski, J. G., Clore, G. M., Schaad, O., Felsenfeld, G., Trainor, C., Appella, E., Stah, S. J. & Gronenborn, A. M. (1993) *Science* **261**, 438–446.
12. Kissinger, C. R., Liu, B., Martin-Blanco, E., Kornberg, T. B. & Pabo, C. O. (1990) *Cell* **63**, 579–590.
13. Hegde, R. S., Grossman, S. R., Laimins, L. A. & Sigler, P. B. (1992) *Nature (London)* **359**, 505–512.
14. Jordan, S. R. & Pabo, C. O. (1988) *Science* **242**, 893–899.
15. Anderson, J. E., Ptashne, M. & Harrison, S. C. (1987) *Nature (London)* **326**, 846–852.
16. Aggarwal, A. K., Rodgers, D. W., Drottar, M., Ptashne, M. & Harrison, S. C. (1988) *Science* **242**, 899–907.
17. Wolberger, C., Dong, Y., Ptashne, M. & Harrison, S. C. (1988) *Nature (London)* **335**, 789–795.
18. Mondragon, A. & Harrison, S. C. (1991) *J. Mol. Biol.* **219**, 321–334.
19. Rodegers, D. W. & Harrison, S. C. (1993) *Structure* **1**, 227–240.
20. Shultz, S. C., Shields, G. C. & Steitz, T. A. (1991) *Science* **253**, 1001–1007.
21. Brennan, R. G., Roderick, S. L., Takeda, Y. & Matthews, B. W. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 8165–8169.
22. Feng, J.-A., Johnson, R.-C. & Dickerson, R. E. (1994) *Science* **263**, 348–355.
23. Clark, M. L., Halay, E. D., Lai, E. & Barley, S. K. (1993) *Nature (London)* **364**, 412–420.
24. Pavletich, N. P. & Pabo, C. O. (1991) *Science* **252**, 809–817.
25. Fairall, L., Schwabe, J., Chapman, L., Finch, J. T. & Rhodes, D. (1993) *Nature (London)* **366**, 483–487.
26. Pavletich, N. P. & Pabo, C. O. (1993) *Science* **261**, 1701–1707.
27. Luisi, B. F., Xu, X. W., Otwinowski, Z., Freedman, L. P., Yamamoto, K. R. & Sigler, P. B. (1991) *Nature (London)* **352**, 497–505.
28. Seeman, N. C., Rosenberg, J. M. & Rich, A. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 804–808.
29. Lehming, N., Sartorius, J., Kisters-Woike, B., von Wilcken-Bergmann, B. & Müller-Hill, B. (1991) in *Nucleic Acids and Molecular Biology*, eds. Eckstein, F. & Lilley, D. M. J. (Springer, Heidelberg), Vol. 5, pp. 114–125.
30. Kisters-Woike, B., Lehming, N., Sartorius, J., von Wilcken-

Bergmann, B. & Müller-Hill, B. (1991) *Eur. J. Biochem.* **198**, 411–419.
31. Desjarlais, J. R. & Berg, J. M. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 2256–2260.
32. Klevit, R. E. (1991) *Science* **253**, 1367–1393.
33. Suckow, M., von Wilcken-Bergmann, B. & Müller-Hill, B. (1993) *EMBO J.* **12**, 1193–1200.
34. Treissman, J., Harris, E., Wilson, D. & Desplan, C. (1992) *BioEssays* **14**, 145–150.
35. Suzuki, M. (1993) *EMBO J.* **12**, 3221–3226.
36. Suzuki, M. (1994) *Structure* **2**, 317–326.
37. Suzuki, M., Gerstein, M. & Yagi, N. (1994) *Nucleic Acids Res.* **22**, 3397–3405.
38. Suzuki, M. (1994) *Proc. Jpn. Acad.* **B70**, 96–99.
39. Suzuki, M. & Chothia, L. (1994) *Proc. Jpn. Acad.* **B70**, 58–61.
40. Suzuki, M. & Yagi, N. (1994) *Proc. Jpn. Acad.* **B70**, 62–66.
41. Deeley, M. & Yanofsky, C. (1992) *J. Bacteriol.* **151**, 942–951.
42. Seiler-Tuyns, A., Walker, P., Martinez, E., Mérillat, A.-M., Givel, F. & Wahli, W. (1986) *Nucleic Acids Res.* **14**, 8755–8770.
43. Thukral, S. K., Eisen, A. & Young, E. T. (1991) *Mol. Cell. Biol.* **11**, 1566–1577.
44. Suzuki, M., Neuhaus, D., Gerstein, M. & Aimoto, S. (1994) *Protein Eng.* **7**, 461–470.
45. Umesono, K., Murakami, K. K., Thompson, C. C. & Evans, R. M. (1991) *Cell* **65**, 1255–1267.

Molecular Cloning of Sequence-Specific DNA Binding Proteins Using Recognition Site Probes

Harinder Singh[1,2], Roger G. Clerc[1] and Jonathan H. LeBowitz[1]

[1]Massachusetts Institute of Technology

[2]University of Chicago

ABSTRACT

Genes encoding sequence-specific DNA binding proteins can be isolated by screening λgt11 expression libraries with recognition site DNAs. This strategy is derived from that developed for the isolation of genes using antibody probes. Many different genes encoding transcriptional regulatory proteins have been cloned using this strategy. The DNA binding domains of these regulatory proteins contain different structural motifs including the helix-turn-helix, the "zinc finger" and the "leucine zipper". Various aspects of the screening strategy are evaluated and a detailed protocol is provided. In addition to binding site DNAs, protein and nucleotide probes have been successfully used to screen expression libraries. Therefore ligand based expression screening may be quite general in scope.

INTRODUCTION

Sequence-specific DNA binding proteins play a central role in deciphering the structural and regulatory information encoded in cellular and viral genomes. They function to initiate as well as control the transcription, replication and site-specific recombination of DNA sequences. Biochemically, these proteins determine the specificity and reactivity of enzymatic assemblies that act on DNA.

In genetically tractable prokaryotes and eukaryotes, most sequence-specific DNA binding proteins have been identified as the products of trans-acting regulatory loci. In many complex eukaryotic organisms a similar approach to their identification has not been possible. Instead, the recent application of sensitive DNA binding assays, in particular, DNase I footprinting (13) and gel electrophoresis of protein-DNA complexes (12,14), has led to the detection and characterization of numerous sequence-specific DNA binding proteins. A majority of these proteins bind selectively to distinct transcriptional control elements and are thereby implicated in regulating the activity of their target genes (31). The isolation of recombinant clones encoding such proteins would facilitate a genetic and biochemical analysis of their structural and functional properties. Prior to the application of the cloning strategy described below, genes encoding sequence-specific DNA binding proteins could be isolated only by screening recombinant DNA libraries with antibody (28,49,50) or oligonucleotide probes (2,25,49): The latter are generated from partial amino acid sequences of the relevant proteins. Both screening strategies are dependent on the availability of substantial amounts of the purified protein. Even though the purification of sequence-specific DNA binding proteins has been greatly facilitated by the development of improved DNA-affinity matrices (4,24, 40), the requirement for very large amounts of starting material (tissue or cells) makes purification on a preparative scale difficult. The new strategy obviates purification of a sequence-specific DNA binding protein for the purpose of isolating its gene. It simply requires an appropriate recombinant DNA library constructed for expression in *Eschericha coli* and a DNA recognition site probe. Therefore, this strategy is ideally suited for isolating clones encoding rare regulatory molecules.

CLONING STRATEGY

The cloning strategy depends on the functional expression in *E. coli* of high levels of the DNA binding domain of a regulatory protein and a strong interaction between this domain and its recognition site. If these conditions are fulfilled, a recombinant clone encoding a sequence-specific DNA binding protein can be detected by probing protein replica filters of an expression library with radiolabeled recognition site DNA. An outline of the steps involved in identifying and analyzing such a clone, using a recombinant library constructed in the expression vector λgt11, is depicted in Figure 1. The initial phase involves the identification of a recombinant clone that is specifically detected with the binding site DNA probe (X) but not with DNA probes

Vol. 7, No. 3 (1989)

that lack the given binding site or contain a mutant version of it (Y), see Figure 1. Such a clone is then shown to encode a β-galactosidase fusion protein of the expected DNA binding specificity. This strategy is derived from that developed for the isolation of genes using antibodies to screen recombinant expression libraries (19,52,53).

Using a ^{32}P-labeled recognition site DNA probe with a specific activity of 10^7 cpm/pmol (*ca.* 10^8 cpm/μg), it is possible to detect 10^{-2} fmol of active protein in a plaque (assuming a 1:1 stoichiometry for the protein-DNA complex). This detection limit represents 1 pg of a β-galactosidase fusion protein (*ca.* M_r 170,000), which is an amount that is well below the expected level of expression for such a protein in a plaque of the desired recombinant λgt11 phage. In fact, overexpression of the *lacZ* fusion gene should result in the accumulation of *ca.* 100 pg of the fusion protein in a phage plaque, assuming that there are 10^5 infected cells/plaque and that the β-galactosidase fusion protein represents 1% of the total protein mass (0.1 pg) of an infected cell. The sensitivity of detection achieved with a ^{32}P-labeled recognition site probe (see above) is comparable to that attained with an ^{125}I-labeled primary antibody (3) or a detection system based on a secondary antibody conjugated with alkaline phosphatase (29). A comparison of the signals generated by a DNA binding site probe and an antibody directed against the corresponding protein is illustrated in Figure 2. The λgt11 recombinant (λEB) encodes a β-galactosidase fusion protein that contains the DNA binding domain of the Epstein-Barr virus nuclear antigen EBNA-1 (38,44). A protein replica filter prepared from a mixed plating of λEB and control λgt11 recombinant phage was screened initially with a recognition site DNA probe (*oriP*) that contains two high affinity binding sites for EBNA-1 and subsequently with antibodies directed against EBNA-1. In this case, the higher signal obtained with the DNA binding site probe was attributed to a less sensitive secondary antibody conjugate containing horseradish peroxidase (29) used in immuno-screening. Note that the patterns of plaques detected by the two types of probes are superimposable. Therefore, a DNA binding site probe can be used to detect a suitable recombinant phage with the same fidelity as an antibody.

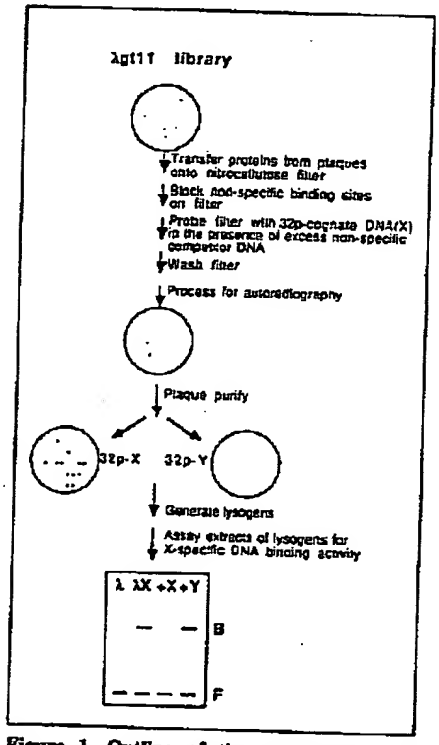

Figure 1. Outline of the strategy for the molecular cloning of sequence-specific DNA binding proteins using the expression vector λgt11. X is a recognition site DNA probe, whereas Y is a control DNA probe that lacks the given recognition site or contains a mutant version of it. The intial phase involves the identification of λgt11 recombinants that are specifically detected with DNA probe X (λX). After plaque purification, the gel electrophoresis DNA binding assay is used to analyze extracts of λX and λgt11 (λ) lysogens. Radiolabeled X-DNA is used as a probe in the binding reactions. F and B refer to free and bound X-DNA, respectively. Reactions in lanes +X and +Y are carried out with the λX extracts and contain an excess of either unlabeled X-DNA or unlabeled Y-DNA as competitors.

SCREENING OF EXPRESSION LIBRARIES

Using screening conditions developed with a model system, Singh et al. (44) isolated a cDNA clone that encodes an enhancer binding protein (H2TF1/NFκB in Table 1). This human cDNA clone was detected by screening a λgt11 expression library with a binding site probe derived from the enhancer of a major histocompatibility complex (MHC) class I gene. The recombinant clone successfully satisfied the criteria depicted in Figure 1. It was detected only with the wild type MHC element (GGGGATTCCCC) probe but not with control DNAs that lack the MHC element or contained a mutant version. Secondly, it specified a β-galactosidase fusion protein which bound specifically to the MHC element in a gel mobility assay. The binding

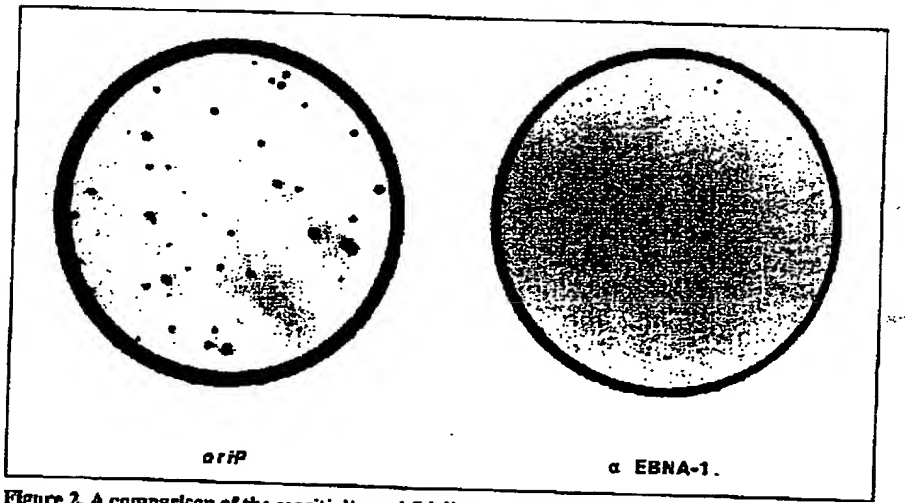

Figure 2. A comparison of the sensitivity and fidelity of detection achieved with DNA binding site and antibody probes. A filter prepared from a mixing plating of λEB (*ca.* 50 pfu) and control λgt11 recombinants (*ca.* 5×10^3 pfu) was screened initially with *oriP* DNA (44) (2×10^6 cpm/ml) and subsequently with α EBNA-1 antibodies (44) (1:500 dilution of rabbit serum, gift of G. Milman, Johns Hopkins University). In the latter screen, the secondary antibody was a conjugate of goat anti-rabbit IgG with horseradish peroxidase.

Table I. List of Clones Encoding Transcriptional Regulatory Proteins, Isolated by Screening λgt11 cDNA Libraries with Recognition Site DNA Probes

Clone	Binding Site	Genes/Genomes That Contain The Binding Site	References
H2TF1/NFκB*	GGGGATTCCCC	MHC I, β2, Igκ, SV40, HIV, IL-2R, β-IFN	44
NF-A2(Oct-2)	ATGCAAAT	IgH, Igκ, H2B, U1, U2, U6, SV40	7,34,46
NF-A1(Oct-1)	ATGCAAAT	IgH, Igκ, H2B, U1, U2, U6, SV40	47
E12	GGCAGGTGG	Igκ	35
XBP	ND	MHC II (Aα)	21
RF-X	CCCCCTAGCAACAG	MHC II (α-DR)	W. Reith et al., 1989[2]
YB-1	GACTAACCGGTTT	MHC II (α-DR)	9
IRF-1	AAGTGA	β-IFN	33
PRDI-BF	GAGAAGTGAAAGTG	β-IFN	T. Maniatis, personal comm.
Pit-1	GATTACATGAATATTCATGA	Prolactin, Growth hormone	23
MLTF	CACGTGACCG	Adenovirus major late transcription unit, Metallothionein γ-fibrinogen	C. Carr and P. Sharp, personal comm.
CREB	TGACGTC	Somatostatin, enkephalin	J. Hoeffler et al., 1988[1]

Sequences in the binding site column generally represent one member of a set of related motifs that the cloned proteins recognize.

*The same clone has been isolated using the PRDII motif of the β-IFN promoter to screen an expression library (T. Maniatis, personal. comm).

[1]Hoeffler, J.P., T.E. Meyer, Y. Yun, J.L. Jameson and J.F. Habener. 1988. Science *242*:1430-1433.

[2]Reith, W., E. Barras, S. Satola, M. Kobr, D. Reinhart, C. Herrero Sanchez and B. Mach. 1989. PNAS (in press).

site was further delineated by methylation interference analysis of the protein-DNA complex. The isolation of this clone validated the various assumptions on which the screening strategy is based. It also provided the impetus for its application in the isolation of other clones encoding sequence-specific DNA binding proteins.

The isolation of a clone encoding a lymphoid-specific octamer binding protein (NF-A2 (Oct-2) in Table 1) (7,46) demonstrated the usefulness of two modifications. In this case, a multi-site DNA probe, consisting of four copies of a 26 bp oligonucleotide containing the octamer motif (ATGCAAAT) was employed. This increased the sensitivity of detection of the relevant recombinant phage (see below). Furthermore, in this screen, sonicated and denatured calf thymus DNA was used as a nonspecific competitor instead of poly (dI-dC)·poly (dI-dC). This substitution reduced the number of inappropriate recombinant phage that were detected (see below).

Vinson et al. (48) have described a third modification in which the nitrocellulose replica filters were subjected to a denaturation/renaturation regimen prior to screening. This treatment enhanced the sensitivity of detection of a phage λ recombinant encoding the enhancer binding protein (C/EBP) (see below). This report also demonstrated enhancement of the detection signal with a multi-site DNA probe.

In the year following the initial application of this strategy, a large number of mammalian cDNA clones encoding distinct sequence-specific DNA binding proteins have been isolated (Table 1). All of these proteins appear to represent transcription factors which regulate the activity of different promoters and enhancer elements (see Table 1). These examples facilitate the evaluation of different aspects of the screening strategy.

Construction of Expression Library

cDNA synthesis and cloning. Successful screening is critically dependent on the frequency with which functional recombinants (in-frame fusions of the DNA binding domain with a bacterial protein segment) are represented in a given cDNA expression library. The cDNA library should be made from mRNA isolated from a cell or tissue source with the highest levels of the desired DNA binding protein. First-strand cDNA synthesis should be carried out using random primers rather than oligo(dT), since the DNA binding domain may be encoded in the amino-terminal part of the desired

protein (5′ end of the corresponding mRNA). Adaptors rather than linkers are preferred for ligating the cDNA inserts to the vector, since they avoid digestion of the cDNA with a restriction enzyme (18,51). It should be noted that most commercially available cDNA expression libraries are constructed using *Eco* RI linkers. Some of these libraries contain a high frequency of partial cDNA inserts that are flanked by natural *Eco* RI sites, indicating inefficient protection of internal sites during their construction (K. LeClair, personal communication). This can result either in the disruption of a cDNA segment encoding a DNA binding domain or in a decrease of the frequency of recombinants containing in-frame fusions of the DNA binding domain with the bacterial protein segment.

Expression vectors. The phage vector λgt11 appears most suitable for expression screening. It offers the advantages of high cloning efficiency, the expression of relatively stable β-galactosidase fusion proteins and a simple means of preparing protein replica filters. Recently, a new bacteriophage λ expression vector (λZAP) has been described which obviates subcloning of cDNA inserts into plasmid vectors for their analysis (42). The presence of multiple cloning sites makes possible the use of "forced cloning" strategies for expression of cDNA inserts from its *lac* promoter. Unlike λgt11, λZAP expresses fusion proteins containing a small amino terminal segment of β-galactosidase. Therefore, the stability of λZAP encoded fusion proteins may be different from their counterparts encoded in λgt11.

Plasmid expression vectors can also be used to detect clones encoding sequence-specific DNA binding proteins. Figure 3 shows that *E. coli* colonies harboring either an EBNA-1 or bacterioiphage λ O protein-expressing plasmid can be specifically detected using the corresponding binding site DNA probes. Even though phage vectors are advantageous for most cloning applications, plasmid vectors could be used to rapidly generate, screen and analyze recombinants encoding mutant DNA binding domains.

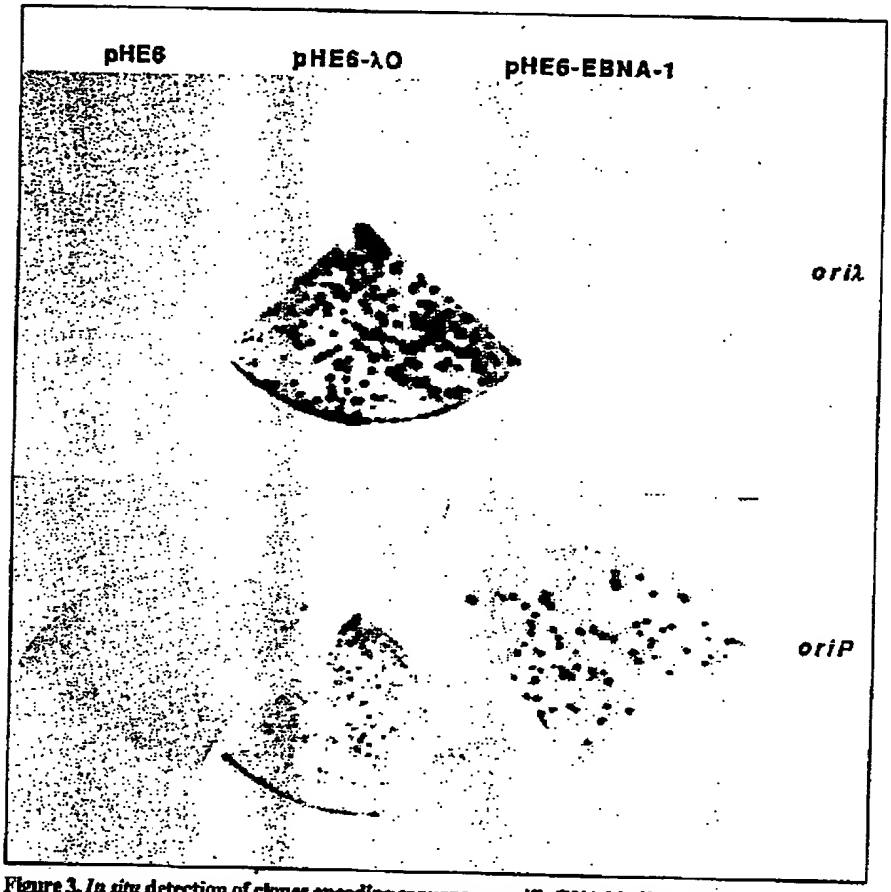

Figure 3. ***In situ*** **detection of clones encoding sequence-specific DNA binding proteins using a plasmid expression system. The recombinant constructs pHE6-EBNA-1 and pHE6-λO were utilized. The former was constructed by fusing in-frame the carboxy-terminal region of EBNA-1 with the amino-terminus of the phage λ N protein in the vector pHE6 (32). This vector contains the λ P_L promoter and a temperature-sensitive λ repressor gene, which permit the thermo-inducible expression of the fusion protein. The second construct, pHE6-λO, allows the inducible expression of the bacteriophage λO protein – a replication initiator which binds specifically to four sites in the λ replication origin, *oriλ*, (39). *E. coli* HB101 colonies harboring the vector or either of its recombinant derivatives were initially grown on nitrocellulose filters at 30° C. The plates were then shifted to 42° C for 2 h in order to induce the high level expression of EBNA-1 or the λO proteins. The filters were transferred to a chromatography tank saturated with chloroform vapors for 15 min (19). This step permitted permeabilization of the cell membranes and adsorption of proteins to nitrocellulose. After permeabilization, the filters were incubated in BLOTTO and then sections were screened with either *oriP* or *oriλ* DNA probes (44) (2×10^6 cpm/ml). The upper row of filter sections was probed with *oriλ* DNA, whereas the lower row was probed with *oriP* DNA. The left, central and right columns represent screens of filter sections prepared from pHE6, pHE6-λO and pHE6-EBNA-1 colonies, respectively.**

Preparation of Protein Replica Filters

Protein replica filters suitable for screening with DNA recognition site probes are most easily prepared using a series of steps derived from the immuno-screening protocol (22, see accompanying protocol). This simple procedure has permitted the detection of many clones encoding different DNA binding proteins, e.g. H2TF1/NFκB, Oct-2, E12, XBP, YB-1, IRF-1, MLTF, CREB (see Table 1). Vinson et al. (48) have shown that processing dried nitrocellulose replica filters through a denaturation/renaturation cycle, using 6 M guanidine hydrochloride, significantly enhances the signal from a λgt11 recombinant encoding C/EBP (see accompanying protocol). However, it is not possible from this report to directly compare the sensitivity of the two protocols in detecting the C/EBP phage, since with the former the replica filters are not dried. The denaturation/renaturation cycle may increase the detection signal by facilitating the correct folding of a

larger fraction of the *E. coli*-expressed protein. Alternatively, it may help to dissociate insoluble aggregates of the fusion protein that form as a consequence of overexpression. This modified procedure has been successfully used to isolate clones encoding Oct-1, Pit-1, PRDI-BF and RF-X (see Table 1). This modification allows the rescreening of the same replica filter with a different DNA probe by repeating the denaturation/renaturation cycle, since the second denaturation step results in dissociation of the DNA probe bound in the first screen.

Screening of Protein Replica Filters

Recognition site DNA probe. The highest affinity site among a set of related sequences should be chosen for the synthesis of an oligonucleotide probe. It has been demonstrated that DNA probes containing a single recognition site can be used to isolate the relevant DNA binding protein clones (H2TF1/NFκB, XBP, YB-1 and MLTF, see Table 1). However, in a number of cases (Oct-2, Oct-1, E12, see Table 1), the signal was appreciably enhanced with DNA probes containing several copies of the appropriate binding site. This effect is demonstrated in Figure 4 with the recombinant phage encoding H2TF1/NFκB (λh3). In this case, the multi-site probe (trimer) was generated by cloning three tandem copies of a 25 bp long oligonucleotide containing the H2TF1/NFκB binding site (GGGGATTCCCC). When equivalent protein replica filters are screened with either the 1-mer (monomer) or the 3-mer (trimer) probe (each end-labeled with ^{32}P to the same specific activity), the latter generates a 3-5 fold higher signal. Multi-site probes can also be prepared for screening simply by catenation of a binding site oligonucleotide with DNA ligase, followed by "nick translation" (48). Such a probe was used to isolate the cDNA encoding Pit-1 (23).

Enhancement of the signal with a multi-site probe may be due to the fact that such a probe can simultaneously interact with two or more immobilized protein molecules, thereby increasing the overall stability of the protein-DNA complexes (see below). This type of DNA probe is particularly suitable for the isolation of clones encoding DNA binding proteins with low affinity for their recognition sites. Given a number of examples in which a multi-site DNA increased the detection signal, it is clearly the preferred type of probe.

Nonspecific competitor DNA. The addition of an excess of nonspecific competitor DNA in the probe solution reduces background as well as minimizes the detection of recombinant phage encoding nonsequence-specific DNA binding proteins. Several different competitor DNAs have been used to successfully screen expression libraries (33,44,46). Screens of such libraries with poly(dI-dC)·poly(dI-dC) as the nonspecific competitor DNA yielded some recombinant phage that encoded proteins which preferentially bind single-stranded DNA (44). As shown in Figure 5, the signal from such phage (e.g., λh1), but not from phage-encoding, sequence-specific DNA binding proteins (e.g., λh3) which encode H2TF1/NFκB), could be efficiently blocked with sonicated and denatured calf thymus DNA at a concentration of 5 µg/ml. The latter DNA further reduced the background signal from the filters. Based on the results of Figure 5 and given that several clones encoding sequence-specific DNA binding proteins have been successfully isolated using sonicated and denatured calf thymus DNA (e.g., Oct-2, MLTF and E12, see Table 1), this nonspecific competitor DNA is preferred.

Binding and wash conditions. The equilibrium association constants of sequence-specific DNA binding proteins range over many orders of magnitude (10^8 - 10^{12} M^{-1}). Consideration of the equilibrium and kinetic constants of a protein-DNA interaction in solution suggests that successful screening may be restricted to proteins with relatively high binding constants, since only these are likely to form complexes with half-lives long enough to withstand the wash protocol (44). For example, if a regulatory protein has an association constant of 10^{10} M^{-1}, then under the screening conditions (the DNA probe is in excess and at a concentration of *ca.* 10^{-10} M), approximately half of the active molecules on the filter will have DNA bound. Since the filters are subsequently washed for 30 min, the fraction of protein-DNA complexes that remain will be determined by their dissociation rate constant. Assuming a diffusion-limited association rate constant of 10^7 M^{-1} S^{-1} (1), the dissociation rate constant in solution will be 10^{-3} S^{-1}. This rate constant translates into a half-life of approximately 10 min. Thus, one-eighth of the protein-DNA complexes should survive the 30 min wash. For a binding constant of 10^9 M^{-1}, about a tenth of

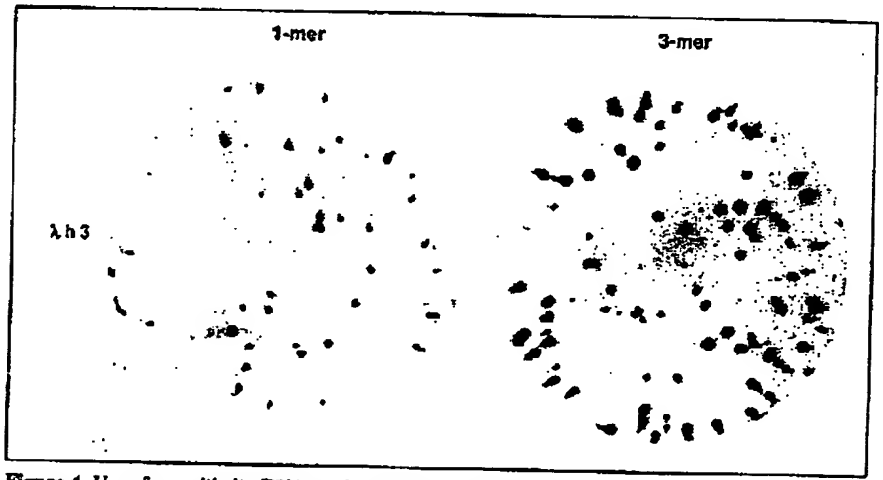

Figure 4. Use of a multi-site DNA probe to enhance signal intensity. Phage λh3 is a λgt11 recombinant that encodes a β-galactosidase fusion protein which binds specifically to an MHC gene regulatory element (44, H2TF1/NFκB in Table 1). Protein replica filters prepared from plating of this phage were screened either with a labeled monomer probe (44, 1-mer) or a labeled multi-site probe containing 3 copies of the same regulatory sequence (3-mer). The latter DNA was a gift of A. Baldwin, Massachusetts Institute of Technology. Both probes were labeled to the same specific activity and each used at a concentration of 2×10^6 cpm/ml.

the active protein molecules will have DNA bound, but virtually all of this signal should be lost since the half-life of these complexes in solution is approximately 1 min. It is unclear whether the equilibrium and kinetic constants of a protein-DNA interaction in solution accurately describe the binding of a DNA probe to a matrix of protein immobilized on a filter. Thus, it may be possible to isolate recombinants encoding proteins with binding constants of 10^9 M^{-1} or lower. The sensitivity of detection of a phage encoding a low affinity variant of the Oct-2 protein is markedly enhanced by using a DNA probe containing multiple binding sites (46). Since the association constants of DNA-binding regulatory proteins are dependent on ionic strength, temperature and pH, these parameters can be manipulated in the binding and wash steps to optimize the detection of a relevant recombinant protein. Finally, if the DNA binding protein being cloned has an exogenous metal ion requirement (e.g., Mg^{2+}), the binding and wash buffers should be appropriately supplemented.

CHARACTERIZATION OF RECOMBINANT DNA BINDING PROTEINS

After the isolation of a recombinant phage that is specifically detected with a given binding site probe, but not with control DNAs, it is necessary to demonstrate that this clone encodes a recombinant protein of the expected DNA binding specificity. In the case of a λgt11 recombinant, this is simply achieved by isolating lysogenized *E. coli* clones and assaying extracts of induced lysogens for a β-galactosidase fusion protein that specifically binds the recognition site probe used in the screen (see accompanying protocol). Chemical and enzymatic footprinting in conjunction with the analysis of mutant binding sites are required to rigorously characterize the DNA binding specificity of the recombinant protein. The criteria used to relate a recombinant protein cloned by this strategy with a previously characterized native protein are discussed in the following section.

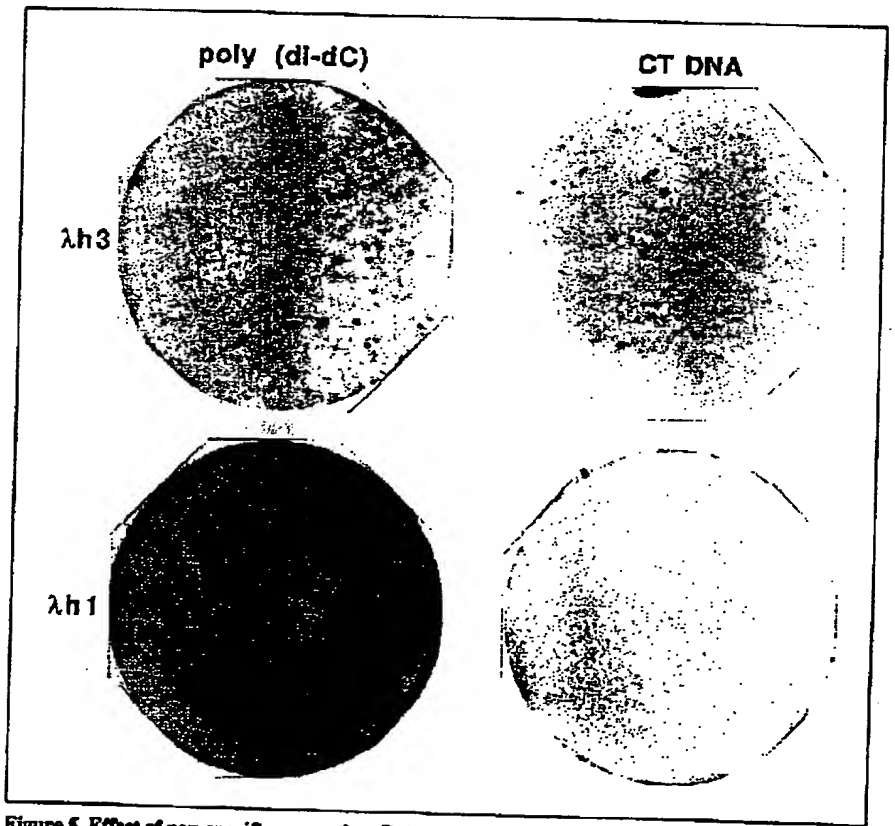

Figure 5. Effect of non-specific competitor DNAs on the detection of recombinant phage encoding different types of DNA binding proteins. λh3 is a recombinant phage that encodes a sequence-specific DNA binding protein (see legend to Figure 4), whereas λh1 encodes a single-stranded DNA binding protein. Protein replica filters prepared from partially purified stocks of λh3 (upper row) and λh1 (lower row) were screened with ^{32}P-labeled MHC 1-mer DNA (see legend to Figure 4, 10^6 cpm/ml) in the presence of either poly(dI-dC)-poly(dI-dC) (left column, 10 μg/ml) or sheared and denatured calf thymus DNA (CT DNA, right column, 5 μg/ml).

DISCUSSION

In this article we have reviewed the development of a new strategy for the molecular cloning of sequence-specific DNA binding proteins. This strategy circumvents purification of such a DNA binding protein for the purpose of isolating its gene. It simply requires a cDNA library constructed in the phage λgt11 and a DNA recognition site probe. As a result of its simplicity and its potential to isolate rare cDNA clones, this strategy is expected to greatly facilitate the analysis of proteins that regulate transcription, DNA replication and site-specific recombination. In fact, within a year of its introduction, more than ten cDNA clones that encode distinct transcriptional regulatory proteins have been isolated using this strategy (see Table 1).

The DNA binding domains of a large number of regulatory proteins contain either a helix-turn-helix motif or the "zinc finger" motif (10,15,36, 41). Clones encoding proteins with either of these structural motifs can be detected by *in situ* screening with the relevant recognition site DNAs. The protein encoded by H2TF1/NFκB cDNA clone contains two "zinc fingers" in its DNA binding domain (Baldwin, LeClair, Singh and Sharp, unpublished results). In contrast, the Oct-2 and Oct-1 cDNA clones encode proteins wtih a predicted helix-turn-helix motif (7,34,46,47). Thus, the screening method appears not to be restricted to a sub-class of DNA binding domains.

Many sequence-specific DNA binding proteins are functional homodimers. The binding sites of these proteins exhibit two-fold rotational symmetry. In these cases the affinity of

Protocol

The following protocol is derived from the one detailed by Singh (44,45) and includes a description of the denaturation/renaturation regimen of Vinson et al. (48).

Materials:

The standard pair of λgt11 host strains, *E. coli* Y1090 and Y1089 are employed (52). Isopropyl-β-D-thiogalactopyranoside (IPTG) is from Pharmacia (Piscataway, NJ). Nitrocellulose filters (82 and 132 mm) type BA85/23 and Elutip-d disposable columns are from Schleicher and Schuell (Keene, NH).

Preparation of radiolabeled recognition site DNA probe:

1. Digest 20 µg (ca. 10 pmol) of a recombinant pUC plasmid DNA, which contains multiple tandem copies (2-10) of the recognition site in the polylinker, to completion with restriction enzymes whose sites flank the insert such as *Hind* III and *EcoR* I. (Probes longer than 200 bp yield higher non-specific signals.)
2. After the plasmid is digested, add the following components to the reaction: 1 M Tris HCl, pH 7.5, to a final concentration of 50 mM; dCTP, dGTP, dTTP to 100 µM each final concentration; 200 µCi of ^{32}P-dATP (5000 Ci/mmol); 10 units *E. coli* DNA polymerase I (Klenow fragment).
3. Incubate at room temperature for 30 min.
4. Add dATP to a final concentration of 100 µM.
5. Continue incubation for an additional 30 min.
6. Stop the end-labeling reaction by adding EDTA to 20 mM.
7. Extract with an equal volume of phenol/chloroform (1:1). Add ammonium acetate, pH 7.5, to a final concentration of 2 M and precipitate with 2.5 volumes of ethanol at -70° C for 10 min.
8. Resuspend the DNA pellet in 200 µl water.
9. Add 22 µl of 3M sodium acetate (pH 7.5) and reprecipitate with 2.5 volumes of ethanol.
10. Resuspend the DNA pellet in sample buffer and separate the labeled DNA fragments by electrophoresis in a 6% non-denaturing polyacrylamide gel.
11. Visualize the labeled fragments by autoradiography (1 min exposure).
12. Cut out a gel slice containing the labeled recognition site DNA fragment. Crush the gel slice in a microfuge tube.
13. Incubate the crushed gel in 1.5 ml elution buffer [20 mM Tris HCl, pH 7.5, 200 mM NaCl, 1 mM EDTA] overnight at 4° C on a rotator.
14. Pellet the crushed acrylamide and load the supernatant on an Elutip-d disposable column. Purify according to the manufacturer's instructions.
15. Determine the total radioactivity incorporated into the probe by scintillation counting and store the probe at 4° C.

Note: These reaction conditions yield DNA probes with specific activities of 2×10^7 to 4×10^7 cpm/pmol. The probe yield should be 10^8 to 2×10^8 cpm. This amount of probe is sufficient to screen twenty 132 mm nitrocellulose filters representing ca. 10^6 plaques (see screening conditions below). Some DNA probe preparations generate a large number of false positives. This problem can be eliminated by filtering the probe solution through a 0.45 µm Gelman acrodisc membrane (T. Hayes, personal communication).

Preparation of the nitrocellulose filter replicas:

1. Grow *E. coli* Y1090 to saturation in LB-medium containing 0.2% maltose and 50 µg/ml ampicillin at 37° C.
2. Take 500 µl aliquots of the Y1090 culture and infect each with $3\text{-}5 \times 10^4$ pfu of the λgt11 cDNA library.
3. Incubate 15 min at 37° C to allow phage adsorption to the cells.
4. In the meantime melt 100 ml top agarose (0.7% agarose in LB-medium) and equilibrate the solution at 47° C.
5. Add 9 ml of top agarose to each 500 µl aliquot of infected cells and invert the mixture twice.
6. Spread each mixture quickly on a prewarmed and dry 150 mm LB/ampicillin plate.
7. Incubate the LB-plates at 42° C until tiny plaques are visible (ca. 3 h).
8. In the meantime soak 132 mm nitrocellulose filters in 10 mM IPTG for 30 min and then air dry them.
9. Overlay each LB-plate from step 7 with a nitrocellulose filter from step 8. Avoid trapping air bubbles between the filter and the top agarose.
10. Incubate the LB-plates at 37° C for 6 h.
11. Cool LB-plates at 4° C for 10 min. Mark position of filter on each plate.

Proceed either to step 12 or 12A (see Application section - Preparation of protein replica filters).

12. Lift nitrocellulose filters and immerse each, protein side up, in a 150 mm petri dish containing 50 ml BLOTTO [5% Carnation non-fat milk powder, 50 mM Tris HCl, pH 7.5, 50 mM NaCl, 1 mM EDTA, 1 mM DTT]. Incubate for 60 min at room temperature with gentle swirling on an orbital platform shaker. Avoid trapping air bubbles while immersing the filters.
13. Transfer each filter to a 150 mm petri dish containing 50 ml binding buffer [10 mM Tris HCl, pH 7.5, 50 mM NaCl, 1 mM EDTA, 1 mM DTT] and incubate as for step 12 for 5 min. Repeat this wash step twice with fresh binding buffer. Filters can be stored in binding buffer at 4° C for up to 24 h prior to screening.

Denaturation/Renaturation Protocol

12A. Lift nitrocellulose filters and air dry them for 15 min at room temperature.

13A. Immerse filters in Hepes binding buffer [25 mM HEPES, pH 7.9, 25 mM NaCl, 5 mM $MgCl_2$, 0.5 mM DTT] supplemented with 6 M guanidine hydrochloride (GuHCl). Incubate with gentle shaking at 4° C for 10 min. All filters can be processed in the same petri dish. Use 100 ml per 15 filters. Repeat this step with fresh Hepes binding buffer containing 6 M GuHCl.

14A. Incubate the filters in Hepes binding buffer containing 3 M GuHCl for 5 min at 4° C (a 1:1 dilution of the 6 M GuHCl solution from the previous step). Repeat this step four times. Each time use Hepes binding buffer that contains a 2-fold dilution of the GuHCl from the previous step.

15A. Incubate the filters in 50 ml of Hepes binding buffer for 5 min at 4° C. Repeat this step and then block the filters by incubating in BLOTTO (see step 13) at 4° C for 30 min.

16A. Immerse filters in Hepes binding buffer supplemented with 0.25% Carnation non-fat milk powder for 1 min at 4° C. Screen filters as below.

Screening of nitrocellulose filter replicas:

1. Incubate each filter in 25 ml binding buffer containing 2.5×10^7 cpm of ^{32}P-labeled DNA recognition site probe and 125 µg denatured, sonicated calf thymus DNA (calf thymus DNA is sheared to *ca.* 1 kb, heat denatured (10 min at 99° C) and quenched on ice before use). Use 150 mm petri dishes for binding incubation and shake gently for 60 min at room temperature. The probe solution can be reused with up to five filters. Filters can also be screened in batches and at 4° C (48).
2. Wash each filter four times (7.5 min each wash, 30 min total) at room temperature with 50 ml aliquots of the binding buffer.
3. Dry filters on blotting paper and perform autoradiography with a tungstate intensifying screen at -70° C for 12 to 24 h.

Identification and purification of sequence specific clones:

1. Identify the presumptive positive phage plaques by aligning the autoradiographs with the LB-plates. To reduce the number of false-positives, generate autoradiography exposures of varying times with the primary filters. Short versus long exposures help to distinguish spots with intense centers (likely to be artifacts) from those with a diffuse halo-like appearance (likely to represent true positives).
2. Isolate agarose plugs corresponding to positive signals and generate secondary phage stocks according to Maniatis et al. (30).
3. Mix a 200 µl aliquot of an overnight *E. coli* Y1090 culture (grown in LB + 0.2% maltose + 50 µg/ml ampicillin) with *ca.* 5×10^3 pfu of the secondary phage stocks (*ca.* titers of secondary phage stocks are $1\text{-}5 \times 10^8$ pfu/ml).
4. Incubate 15 min at 37° C.
5. Add 3 ml of top agarose pre-equilibrated at 47° C.
6. Invert the solution twice and spread on a prewarmed and dry 100 mm LB-plate.
7. Proceed as described in step 7 "Preparation of the nitrocellulose filter replicas" using 82 mm nitrocellulose filters.
8. For screening 82 mm filters, use 10 ml aliquots of binding solution in 100 mm petri dishes.
9. Screen secondary filters representing true positives with the wild-type recognition site DNA probe as well as with control DNA probes that either lack the binding site or contain mutant versions.
10. Plaque purify phage which are specifically detected with the wild-type recognition site probe but not with control DNA probes.

Isolation of recombinant phage lysogens:

1. Grow *E. coli* Y1089 to saturation in LB-medium containing 0.2% maltose and 50 µg/ml ampicillin at 37° C.
2. Dilute the saturated cell culture 100-fold in LB-medium supplemented with 10 mM $MgCl_2$.
3. Mix a 100 µl aliquot of the diluted culture with 5 µl of purified recombinant phage stock (*ca.* 10^{10} pfu/ml) and incubate 20 min at 32° C.
4. Dilute the infected cell suspension 1000 fold in LB-medium and plate 100 µl aliquots of the diluted cell suspension on 100 mm LB-plates containing 50 µg/ml ampicillin.
5. Incubate LB-plates at 32° C overnight. At 32° C the temperature sensitive λgt11 encoded repressor is functional and programs the establishment of the lysogenic state.
6. Test colonies for temperature sensitive growth by picking onto two LB-plates containing 50 µg/ml ampicillin. Incubate one plate at 42° C and the other at 32° C. Clones that grow at 32° C but not at 42° C represent lysogens. Lysogens should arise at a frequency to 10-80%.

Preparation of crude cell extracts from recombinant phage lysogens.

1. Grow overnight cultures of recombinant phage lysogens in LB-medium containing 50 µg/ml ampicillin at 32°C.
2. Mix a 2 ml aliquot of LB-medium/ampicillin with 20 µl of an overnight culture of a lysogen and incubate at 32° C with good aeration.
3. Carefully monitor growth of these cultures and shift to a 44° C incubator for 20 min when the $OD_{600} = 0.5$ (*ca.* 3 h incubation at 32° C). incubator for 60 min.
4. Adjust the culture to 10 mM IPTG to induce the expression of the β-galactosidase fusion protein and shift the culture to a 37° C
5. Spin a 1 ml aliquot of the induced culture in a microfuge for 1 min at room temperature.
6. Discard the supernatant and resuspend the pellet in 100 µl extract buffer [50 mM Tris HCl, pH 7.5, 1 mM EDTA, 1 mM DTT, 1 mM PMSF (freshly prepared)].
7. Quickly freeze the resuspended cells in liquid nitrogen.
8. Thaw the frozen cell suspension, adjust to 0.5 mg/ml lysozyme and incubate for 15 min on ice.
9. Adjust the cell suspension to 1 M NaCl, mix thoroughly and incubate for 15 min on a rotator at 4° C.
10. Centrifuge the lysates in microfuge for 30 min at 4° C.
11. Dialyze the supernatant on Millipore filter (type VS, 0.025 µm pore size) against 100 ml extract buffer for 60 min at 4° C. Millipore filters should be floated on dialysis buffer in a 150 mm petri dish before applying samples.
12. Freeze the dialyzed extract immediately and store at -70° C till needed.
13. The DNA binding properties of the fusion protein in the extract can be tested in various ways including the gel mobility shift assay (5,44) and DNase I footprinting (13,25).

Note: Some β-galactosidase fusion proteins are poorly solubilized by the above extraction procedure. In these cases detergents and/or denaturants may be required to effect solubilization (C. Carr, personal communication).

the monomer for the complete binding site is significantly lower than that of the dimer (37,41). Clones encoding such homodimeric proteins can also be detected by *in situ* screening. The bacteriophage λ O protein appears to bind its dyad symmetric recognition site in *ori*λ DNA as a dimer (Roberts and McMacken, personal communication). A clone encoding this protein can be specifically detected by *in situ* screening of bacterial colonies using *ori*λ DNA as probe (see Figure 3). The mammalian protein C/EBP also appears to require dimerization for sequence-specific binding (Landschulz, Johnson and McKnight, personal communication). A λgt11 recombinant encoding this protein can be detected by screening plaque lifts with the corresponding DNA binding site probe (48). Interestingly, the region of C/EBP required for dimerization, the "leucine zipper," is shared by a number of regulatory proteins including GCN4, Fos, Myc and Jun (28). Recently, Murre et al. (35) have used the screening strategy described herein to isolate cDNAs encoding a mammalian enhancer binding protein (E12, Table 1) that requires a new type of dimerization domain for DNA binding. These examples clearly show that clones encoding proteins that bind DNA as homodimers, using different dimerization domains, can be successfully screened as a consequence of their functional expression in *E. coli*.

Most functional DNA binding domains, including elements required for dimerization, are contained within relatively small protein segments (approximately 60-200 amino-acids, e.g., the DNA binding domains of EBNA-1(38), GAL-4 (26), GCN-4 (20), Spl (25)); therefore, successful screening is not dependent on full-length cDNA clones. It simply requires that a given expression library contain partial cDNA clones spanning the DNA binding domain of the desired protein.

The screening strategy, although a very powerful tool, has limitations. Since it relies on functional expression of a DNA binding domain in *E. coli*, it is highly unlikely to enable the cloning of proteins, which depend either on a cell-specific post-translational modification or a second distinct subunit for high affinity DNA binding. In the case of heterodimeric proteins (6,16), one of the two subunits may bind the recognition site with an affinity that makes the isolation of its gene by *in situ* screening feasible. For example, in the AP-1 and c-Fos complex, c-Fos confers high affinity binding, but the AP-1 subunit alone binds the same recognition site with detectable affinity (8,17). Given the clone for one subunit of a heterodimeric DNA binding protein, it may be possible to clone the gene encoding the second subunit by using a variation of the expression screening approach (see below). Another limitation of this strategy is, initially a recombinant protein can only be related to a previously identified native protein by comparing the DNA binding specificities of the two. However, in a situation where multiple DNA binding proteins recognize the same sequence, this criterion is very difficult to apply (21,44). Eventually, direct structural analyses are necessary to resolve this issue. Antibodies generated against the cloned protein permit the detection of shared antigenic determinants (47). Peptide mapping performed on analytical amounts of the native and cloned proteins constitutes a definitive structural comparison (7). A third limitation of this strategy is that its application can result in the isolation of recombinant phage whose β-galactosidase fusion proteins do not appear to bind DNA with detectable affinity in solution (T. Kristie, personal communication).

The strategy of cloning a gene on the basis of detection of its functional recombinant product with a ligand probe, has considerable potential. It may be possible to use different types of ligands, including RNA recognition sites, hormones, protein subunits (e.g., a subunit of a heterodimeric DNA binding protein), nucleotides, metal ions etc., to directly clone genes that encode the relevant proteins. During the development and application of the strategy reviewed in this article, a few ligand-mediated screens of this type have been described. The cDNA for a calmodulin-binding protein has been cloned using iodinated calmodulin as a probe to screen a λgt11 expression library (43). Similarly, λgt11 clones expressing the regulatory subunit of a cAMP dependent protein kinase could be detected by an *in situ* screen of a library using ^{32}P-labeled cAMP as a probe (27). Finally, mutants of *ras* that are defective in GTP binding have been isolated by using ^{32}P-labeled GTP in an *in situ* colony-binding assay (11). Thus, the principle underlying this strategy appears quite general in its scope.

ACKNOWLEDGMENTS

The experiments described in this overview were carried out in the laboratory of Dr. Phillip A. Sharp. We gratefully acknowledge Phillip Sharp for many stimulating discussions and for his critical comments on this manuscript. We also thank our colleagues at the Center for Cancer Research for many helpful suggestions, in particular A. Baldwin, M. Brown, M. Garcia-Blanco, M. Griot, T. Hayes, T. Kristie and K. LeClair.

H.S. acknowledges postdoctoral support from the Jane Coffin Childs Fund. R.G.C. is the recipient of a Ciba-Geigy, Basel (Switzerland) fellowship and a Schweizerischer National Funds stipend. J.H.L. is a Special Fellow of the Leukemia Society of America.

REFERENCES

1. Berg, O.G., R.B. Winter and P.H. von Hippel. 1982. How do genome-regulatory proteins locate their DNA target sites? Trends in Biochem. Sci. *7*:52-55.
2. Bodner, M., J.L. Castrillo, L.E. Theill, T. Deerinck, M. Ellisman and M. Karin. 1988. The pituitary-specific transcription factor GHP-1 is a homeo box containing protein. Cell *55*:505-518.
3. Broome, S. and W. Gilbert. 1978. Immunological screening method to detect specific translation products. Proc. Natl. Acad. Sci. USA *75*:2746-2749.
4. Chodosh, L.A., R.W. Carthew and P.A. Sharp. 1986. A single polypeptide possesses the binding and transcription activities of the Adenovirus major late transcription factor. Mol. Cell. Biol. *6*:4723-4733.
5. Chodosh, L.A. 1988. Mobility shift DNA-binding assay using gel electrophoresis, 12.2. *In* M. Ausubel, R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith and K. Struhl (Eds.). Current Protocols in Molecular Biology. John Wiley and Sons, New York, NY.
6. Chodosh, L.A., A.S. Baldwin, R.W. Carthew and P.A. Sharp. 1988. Human CCAAT-binding proteins have heterologous subunits. Cell *53*:11-24.
7. Clerc, R.G., L.M. Corcoran, J.H. LeBowitz, D. Baltimore and P.A. Sharp. 1988. The B-

cell specific Oct-2 protein contains Pou box and homeo box-type domains. Genes and Development *2*:1570-1581.
8. Curran, T. and B.R. Franza. 1988. Fos and Jun: The AP-1 connection. Cell *55*:395-397.
9. Didier, D.K., J. Schiffenbauer, S.L. Woulfe, M. Zacheis and B.D. Schwartz. 1988. Characterization of the cDNA encoding a protein binding to the major histocompatibility complex class II Y box. Proc. Natl. Acad. Sci. USA *85*:7322-7326.
10. Evans, R.M. and S.M. Hollenberg. 1988. Zinc fingers: Guilt by association. Cell *52*:1-3.
11. Feig, L.A., B.T. Pan, T.M. Roberts and G.M. Cooper. 1986. Isolation of *ras* GTP-binding mutants using an *in situ* colony-binding assay. Proc. Natl. Acad. Sci. USA *83*:4607-4611.
12. Fried, M. and D. Crothers. 1981. Equilibrium and kinetics of *lac* repressor-operator interactions by polyacrylamide gel electrophoresis. Nuc. Acids Res. *9*:6505-6525.
13. Galas, D. and A. Schmitz. 1978. DNase footprinting: A simple method for the detection of protein-DNA binding specificity. Nuc. Acids Res. *5*:3157-3170.
14. Garner, M. and A. Revzin. 1981. A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: Applications to components of the *E. coli* lactose operon regulator system. Nuc. Acids Res. *9*:3047-3060.
15. Gehring, W.J. 1987. Homeo boxes in the study of development. Science *236*:1245-1252.
16. Hahn, S. and L. Guarente. 1988. Yeast Hap2 and Hap3 - Transcriptional activators in a heterodimeric complex. Science *240*:317-321.
17. Halazonetis, T.D., K. Georgopoulos, M.E. Greenberg and P. Leder. 1988. c-Jun dimerizes with itself and with c-Fos, forming complexes of different DNA binding affinities. Cell *55*:917-924.
18. Haymerle, H., J. Herz, G.M. Bressan, R. Frank and K.K. Stanley. 1986. Efficient construction of cDNA libraries in plasmid expression vectors using an adaptor strategy. Nuc. Acids Res. *21*:8615-8624.
19. Helfman, D.M., J.R. Feramisco, J.C. Fiddes, G.P. Thomas and S.H. Hughes. 1983. Identification of clones that encode chicken tropomyosin by direct immunological screening of a cDNA expression library. Proc. Natl. Acad. Sci. USA *80*:31-35.
20. Hope, I.A. and K. Struhl. 1986. Functional dissection of a eukaryotic transcriptional activator protein, GCN4 of yeast. Cell *46*:885-894.
21. Hsiou-Chi, L., M.R. Boothby and L.H. Glimcher. 1988. Distinct cloned class II MHC DNA binding proteins recognize the X box transcription element. Science *242*:69-71.
22. Huynh, T.V., R.A. Young and R.W. Davis. 1985. Construction and screening cDNA libraries in λgt10 and λgt11, p. 49-78. *In* D.M. Glover (Ed.), DNA Cloning, Vol. 1: A Practical Approach. IRL Press, Oxford.
23. Ingraham, H.A., R. Chen, H.J. Mangalam, H.P. Elsholtz, S.E. Flynn, C.R. Lin, D.M. Simmons, L. Swanson and M.G. Rosenfeld. 1988. A tissue-specific transcription factor containing a homeo domain specifies a pituitary phenotype. Cell *55*:519-529.
24. Kadonaga, J.T. and R. Tjian. 1986. Affinity purification of sequence-specific DNA binding proteins. Proc. Natl. Acad. Sci. USA *83*:5889-5893.
25. Kadonaga, J.T., K.R. Carner, F.R. Masiarz and R. Tjian. 1987. Isolation of cDNA encoding transcription factor Sp1 and functional analysis of the DNA binding domain. Cell *51*:1079-1090.
26. Keegan, L., G. Gill and M. Ptashne. 1986. Separation of DNA binding from the transcription-activating function of a eukaryotic regulatory protein. Science *231*:699-704.
27. Lacombe, M.L., D. Ladant, R. Mutzel and M. Veron. 1987. Gene isolation by direct *in situ* cAMP binding. Gene *58*:29-36.
28. Landschulz, W.H., P.F. Johnson, E.Y. Adashi, B.J. Graves and S.L. McKnight. 1988. Isolation of a recombinant copy of the gene encoding C/EBP. Genes and Development *2*:786-800.
29. Leary, J.J., D.J. Brigati and D.C. Ward. 1983. Rapid and sensitive colorimetric method for visualizing biotin-labeled DNA probes hybridized to DNA or RNA immobilized on nitrocellulose: Bio-blots. Proc. Natl. Acad. Sci. USA *80*:4045-4049.
30. Maniatis, T., E.F. Fritsch and J. Sambrook. 1982. Molecular Cloing: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
31. Maniatis, T., S. Goodbourn and J.A. Fischer. 1987. Regulation of inducible and tissue specific gene expression. Science *236*:1237-1245.
32. Milman, G., A.L. Scott, M.S. Cho, S.C. Hartman, D.K. Ades, G.S. Hayward, P.F. Ki, J.T. August and S.D. Hayward. 1985. Carboxyl-terminal domain of the Epstein-Barr virus nuclear antigen is highly immunogenic in man. Proc. Natl. Acad. Sci. USA *82*:6300-6304.
33. Miyamoto, M., T. Fujita, Y. Kimura, M. Maruyama, H. Harada, Y. Sudo, T. Miyata and T. Taniguchi. 1988. Regulated expression of a gene encoding a nuclear factor, IRF-1, that specifically binds to IFN-β gene regulatory elements. Cell *54*:903-913.
34. Müller, M.M., S. Ruppert, W. Schaffner and P. Matthias. 1988. A cloned octamer transcription factor stimulates transcription from lymphoid specific promoters in non-B cells. Nature *336*:544-551.
35. Murre, C., P. Schonleber-McCaw and D. Baltimore. 1989. A new DNA binding and dimerization motif in immunoglobulin enhancer binding, *daughterless*, MyoD and myc proteins. Cell, in press.
36. Pabo, C.O. and R.T. Sauer. 1984. Protein-DNA recognition. Ann. Rev. Biochem. *53*: 293-321.
37. Ptashne, M. 1986. A Genetic Switch. Cell Press and Blackwell Scientific Publications, Cambridge, MA.
38. Rawlins, D.R., G. Milman, S.D. Hayward and G.S. Hayward. 1985. Sequence-specific DNA binding of the Epstein-Barr virus nuclear antigen (EBNA-1) to clustered sites in the plasmid maintenance region. Cell *42*:859-868.
39. Roberts, J.D. and R. McMacken. 1983. The bacteriophage λ O replication protein: Isolation and characterization of the amplified initiator. Nuc. Acids Res. *11*:7435-7452.
40. Rosenfeld, P.J. and T.J. Kelly. 1986. Purification of nuclear factor I by DNA recognition site affinity chromatography. J. Biol. Chem. *261*:1398-1408.
41. Schleif, R. 1988. DNA binding by proteins. Science *241*:1182-1187.
42. Short, J.M., J.M. Fernandez, J.A. Sorge and W.D. Huse. 1988. λZAP: A bacteriophage λ expression vector with *in vitro* excision properties. Nuc. Acids Res. *16*:7583-7600.
43. Sikela, J.M. and W.E. Hahn. 1987. Screening an expression library with a ligand probe: Isolation and sequence of a cDNA corresponding to a brain calmodulin-binding protein. Proc. Natl. Acad. Sci. USA *84*:3038-3042.
44. Singh, H., J.H. LeBowitz, A.S. Baldwin and P.A. Sharp. 1988. Molecular cloning of an enhancer binding protein: Isolation by screening of an expression library with a recognition site DNA. Cell *52*:415-423.
45. Singh, H. 1988. Detection, purification and characterization of cDNA clones encoding DNA-binding proteins, 12.7. *In* M. Ausubel, R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith and K. Struhl (Eds.). Current Protocols in Molecular Biology. John Wiley and Sons, New York, NY.
46. Staudt, L.M., R.G. Clerc, H. Singh, J.H. LeBowitz, P.A. Sharp and D. Baltimore. 1988. Cloning of a lymphoid specific cDNA encoding a protein binding the regulatory octamer DNA motif. Science *241*:577-580.
47. Sturm, R.A., G. Das and W. Herr. 1988. The ubiquitous octamer protein Oct-1 contains a Pou domain with a homeo subdomain. Genes and Develpment, in press.
48. Vinson, C.R., K.L. LaMarco, P.F. Johnson, W.H. Landschulz and S.L. McKnight. 1988. *In situ* detection of sequence-specific DNA binding activity specified by a recombinant bacteriophage. Genes and Development *2*:801-806.
49. Walter, P., S. Green, G. Green, A. Krust, J.-M. Bornert, J.-M. Jeltsch, A. Staub, E. Jensen, G. Scrace, M. Waterfield and P. Chambon. 1985. Cloning of the human estrogen receptor cDNA. Proc. Natl. Acad. Sci. USA *82*:7889-7893.
50. Weinberger, C., S.M. Hollenberg, E.S. Ong, J.M. Harmon, S.T. Brower, J. Cidlowski, E.B. Thompson, M.G. Rosenfeld, R.M. Evans. 1985. Identification of human glucocorticoid receptor complementary DNA clones by epitope selection. Science *228*:740-742.
51. Wu, R., T. Wu and A. Ray. 1987. Adaptors, linkers and methylation. Meth. Enzymol. *152*:343-349.
52. Young, R.A. and R.W. Davis. 1983. Efficient isolation of genes by using antibody probes. Proc. Natl. Acad. Sci. USA *80*:1194-1198.
53. Young, R.A. and R.W. Davis. 1983. Yeast RNA polymerase II genes: Isolation with antibody probes. Science *222*:778-782.

Address correspondence to:

Harinder Singh, Ph.D.
Howard Hughes Medical Institute and
Dept. of Molecular Genetics & Cell Biology
University of Chicago
Chicago, IL 60637